A Study on Efficient and Secure Set Similarity Joins

著者	Mateus SilqueiraHicksonCruz
発行年	2018
その他のタイトル	効率的で安全な集合間類似結合に関する研究
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2017
報告番号	12102甲第8523号
URL	http://hdl.handle.net/2241/00152777

氏　　名	Mateus Silqueira Hickson Cruz			
学位の種類	博　士（　工学　）			
学位記番号	博　甲　第　８５２３　号			
学位授与年月日	平成　３０年　３月　２３日			
学位授与の要件	学位規則第４条第１項該当			
審査研究科	システム情報工学研究科			
学位論文題目	A Study on Efficient and Secure Set Similarity Joins （効率的で安全な集合間類似結合に関する研究）			
主　　査	筑波大学　教授	理学博士	北川　博之	
副　　査	筑波大学　教授	博士（工学）	天笠 俊之	
副　　査	筑波大学　教授	博士（工学）	佐久間　淳	
副　　査	筑波大学　教授	博士（理学）	高橋　大介	
副　　査	産業技術大学院大学　教授	博士（情報理工学）	中野　美由紀	

論　文　の　要　旨

実世界では、文書や画像など、多くのデータが単語あるいは特徴量などを要素とした集合として表現される。このとき、集合として表現されたデータ（レコード）群から類似した全てのペアを列挙する処理は、「集合間類似結合（set similarity join)」と呼ばれ、データクリーニングや文書（画像）マッチングなどに共通に現れる汎用的な処理である。このとき、一つのレコードは一般に大量（数百から数千）の要素で表現され、多くのレコードを扱う際には、マッチング処理の計算コストおよび空間コストが問題となる。さらに近年では個人情報保護やセキュリティに対する要請から、レコードの内容を外部に明かすことなくセキュアに集合間類似結合を実行したいという要求も高まっている。

本論文では、集合間類似結合の高速化ならびにセキュアな集合間類似結合プロトコルの研究に取り組んでいる。まず高速化については、近年注目されている GPU（graphics processing unit）を活用した新たな高速化手法を提案している。GPU では多くのプロセッサを利用した並列度の高い処理が可能な反面、搭載されているメモリ（デバイスメモリ）は高々数 GB であり、大規模なデータをそのまま処理することは難しい。本論文では、集合データを圧縮した上で類似度判定を近似的に行うことができる MinHash 法を適用した新たなアルゴリズムを提案している。実験による性能評価により、精度をほとんど損なうことなく、シングルスレッドの CPU に対して百倍以上の高速化を達成している。

セキュアな集合間類似結合については、完全準同型暗号を利用した新たなプロトコルを提案している。既存の研究では、決定論的暗号アルゴリズムを利用することで、暗号文同士の比較によって集合要素比較を行なっていた。この方法では処理を高速に行なうことができる一方、暗号文から平文を推測される危険性がある。本論文ではこれを避けるため、完全準同型暗号を利用して集合要素の同一性判定を行なうことで安全性を向上している。この際、完全準同型暗号は処理が極めて遅いため、複数

の暗号文を一つの暗号文にまとめる、あるいは各種フィルタリング技術を導入することで処理の高速化を図っている。実データを使った実験を行ない、提案手法の有効性を示している。

審　査　の　要　旨

【批評】

集合間類似結合は、文書を対象としたクリーニングや曖昧性解消、類似画像のマッチングなど、多くの応用で利用されている実用性の高い処理である。大規模データに対する処理の要求やセキュリティに対する社会的な要請などを受け、集合間類似結合を大規模なデータに適用する際の高速化、ならびにデータの内容を外部に明かすことなくマッチングを実行することに対するニーズが高まっている。これを背景に、本研究では、集合間類似結合の高度化とセキュアな集合間類似結合処理に取り組んでいる。まず処理の高速化では、近年注目されているデバイスである GPU（graphics processing unit）を活用した手法を提案している。GPU は処理の高速化が可能である反面、メモリ空間がホストコンピュータから独立しており容量も数 GB しかないため、大規模データの処理には工夫が必要である。そこで MinHash 法を適用し、これを GPU 上で高速に実行する手法を提案し、その有効性を実験的に検証している。その結果、CPU の単一スレッドに対して 100 倍以上の高速化を達成した。また、セキュリティの観点から、完全準同型暗号を利用した新たなプロトコルを提案している。本手法では確率的暗号化を利用しているため、従来法に比べてより安全性が高いという特徴がある。ただし、完全準同型暗号は実行が極めて遅いという課題に対応するため、各種フィルタリングの手法を取り入れるなど、実行効率の改善を図っている。実データを使った実験により提案手法の有効性が示されている。

以上のように、本研究では集合間類似結合に対して、GPU を用いた高速化ならびに完全準同型暗号を用いた安全性の高い処理の新たな手法を提案し、それらの有効性を実験によって示している点において、情報工学上の大きな貢献が認められる。今後は、より大規模なデータに対する対応や手法の精度の理論解析などが期待される。

【最終試験の結果】

平成 30 年 2 月 5 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。